HELLO KITTY & BUBBLEGUM PRINCESS
SALAD FESTIVAL!
SO MANY GREENS, SO MANY POSSIBILITIES!
BUILD YOUR PERFECT SALAD!
In a ReLU network, each layer applies a linear map then a ReLU:
$h^\ell = \mathrm{ReLU}(W^\ell h^{\ell-1} + b^\ell)$
Some information gets zeroed out.
Observability can only stay the same or decrease with depth.
$\mathrm{Rank}(h^\ell) \le \mathrm{Rank}(h^{\ell-1})$
Input
Layer 1
Layer 2
Layer 3
Observability Decay!
OBSERVABILITY IS IMPORTANT!
• Don't lose what you might need later!
• Context can help keep it alive.
• Bypass. Preserve. Enrich.
But if context from the global input is available at deeper layers, it can bypass or reduce the decay!
Use context C from the input x.
Augmented input at layer ℓ:
$\tilde{h}^\ell = [h^{\ell-1}; \phi_\ell(C)]$
The network can recover or preserve information that ReLU would have otherwise lost.
Context is like extra dressing for the salad of computation!
SALAD CHEF
DRESSING BAR
RANCH (robust)
BALSAMIC (deep)
CITRUS (fresh)
SESAME (nutty)
CAESAR (classic)
YOGURT (light)
HONEY MUSTARD (sweet)
AVOCADO (sweet)
AVOCADO CREAMY (smooth)
MAGICAL RAINBOW (extra context!)
TODAY'S SPECIAL: MAXIMUM FLAVOR MINIMUM REGRET!
GOOD SALAD GOOD FRIENDS GOOD CONTEXT = BETTER NETWORKS

Per-layer context in ReLU neural networks loses information about the input directly as the data channel loses it. You can try augmenting with context directly from the input or use Atlas LSH neural networks.